Supporting Information for "Spatial partitioning improves the reliability of biochemical signaling"

Andrew Mugler[1], Filipe Tostevin[1], and Pieter Rein ten Wolde[1]

[1]FOM Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands

This document contains the methods, supporting analytic results, and supplementary Figures S1-S8.

1 Methods

1.1 Spectral solution of the master equation

The chemical master equation (CME) is solved using the method of spectral expansion [1, 2], described in detail in the next section. Briefly, the structure of of the CME, in which the dynamics can be separated into two operators that act only on m or n but not both, allows for its solution to be written in the form $p_{mn}(t) = \sum_{j=0}^{M}\sum_{k=0}^{N} G_{jk}(t;\bar{\beta})\phi_m^j(\alpha)\phi_n^k(\bar{\beta})$, where $\phi_m^j(\alpha)$ is the j^{th} eigenvector of the operator $\mathcal{L}_m(\alpha)$ and similarly for $\phi_n^k(\bar{\beta})$, and $\bar{\beta}$ is an expansion parameter on which p_{mn} does not ultimately depend. The expansion coefficients $G_{jk}(t;\bar{\beta})$ can be calculated straightforwardly, as shown in the next section. Importantly, this spectral expansion dramatically decreases the computational complexity of calculating p_{mn}: rather than solving the $(M+1)(N+1)\times(M+1)(N+1)$ system of the original CME, it is only necessary to solve N linear systems of size $(M+1)\times(M+1)$ for the vectors of coefficients $\vec{G}_k$. We emphasize that since the system has a finite state-space, no approximations are made in using the spectral expansion, and the solution remains exact. Furthermore, the moments of the steady-state distribution p_{mn} can be conveniently expressed in terms of the expansion coefficients G_{jk}; in particular, $\langle n\rangle = G_{01}$ and $\langle n^2\rangle = 2G_{02} + G_{01}$.

1.2 Mutual information

The mutual information between network input and response is given by the standard expression [3] $I[\alpha, m] = \langle\log\{p(\alpha,n)/[p(\alpha)p(n)]\}\rangle$, where the average is taken over the joint distribution $p(\alpha,n) = p(n|\alpha)p(\alpha)$, and $p(n|\alpha) = \sum_{m=0}^{M} p(m,n|\alpha)$ is given by the steady state of the CME. The calculation of the mutual information requires specification of the distribution of input signals $p(\alpha)$. We choose N_α values of α such that $q = \alpha/(\alpha+1) = \langle m\rangle/M$ is uniformly-spaced over the range $0 \le q \le 1$; then $p(n) = \sum_{i=1}^{N_\alpha} p(n|\alpha_i)p(\alpha_i)$ and $p(\alpha_i) = N_\alpha^{-1}$. However, our conclusions are unaffected if we instead take a input distribution that is unimodal or bimodal (Fig. S7). We take $N_\alpha > 30$, for which $I[\alpha, m]$ converges to within 1% of its large-N_α limit (Fig. S8).

1.3 Spatial simulations

The diffusion and reactions of M $\mathcal{X}$ molecules and N $\mathcal{Y}$ molecules are simulated on a two-dimensional square lattice of side length λ using a fixed-time-step integration scheme. During each step of duration δt, each particle is moved to a random neighboring lattice site with probability $p_D = (D/\ell^2)\delta t$, where D is the diffusion constant, and ℓ is the lattice spacing. Molecules have steric interactions on the lattice, such that only one molecule can be present at each lattice site at any time. Attempted moves to an occupied site are rejected, with the particle remaining at its original position. If a molecule in the X^* state is adjacent to a molecule in the Y state, the latter is converted to the Y^* state with probability $p_r = \gamma(\beta\lambda^2/M)\delta t$. To make π partitions, linear diffusion barriers are placed at $i\lambda/\sqrt{\pi}$ in each direction, where $i \in \{0, 1, \ldots, \sqrt{\pi} - 1\}$. A diffusion step which crosses such a barrier is accepted with probability reduced by a factor p_{hop}. The time step δt is chosen sufficiently small that no probability exceeds one.

2 Solution of the master equation by spectral expansion

This section describes the solution via the method of spectral expansion, or the 'spectral method', of the CME introduced in the main text. The spectral method has been used fruitfully in the context of gene regulation to solve CMEs describing cascades [1], bursts [2], and oscillations [4], and a pedagogical treatment is available in [5]. Here we apply the spectral method to coupled reversible switching.

From Eqns. 1-2 of the main text, the stochastic dynamics of the system under study are given by the CME

$$\dot{p}_{mn} = -\left[\mathcal{L}_m(\alpha, M) + \gamma\mathcal{L}_n(\beta_m, N)\right] p_{mn}, \tag{1}$$

where both operators $\mathcal{L}_m$ and $\mathcal{L}_n$ have the form

$$\mathcal{L}_m(\alpha, M) = \alpha\left[1 - \mathbb{E}_m^{-1}\right](M - m) + \left[1 - \mathbb{E}_m^{+1}\right] m, \tag{2}$$

with $\mathbb{E}_m^i f(m) = f(m + i)$ defining the step operator. The CME describes the evolution of the probability of having m $\mathcal{X}$ proteins in the active state and n $\mathcal{Y}$ proteins in the active state, with β_m the coupling function by which $\mathcal{X}$ drives the activation of $\mathcal{Y}$.

2.1 The moments do not close

We first demonstrate that direct computation of the moments from the CME is not possible because the moments do not close. The reason is that a nonlinearity is present in the first term of Eqn. 2 in the form of the factor $\beta_m n$. As a result, the first moment depends on a higher moment, which in turn depends on an even higher moment, and so on.

To see explicitly that the moments do not close, we consider computing the dynamics of the first moment of the driven species, the mean $\langle n\rangle$, by summing the CME over m and n against n. We obtain

$$\frac{1}{\gamma}\partial_t\langle n\rangle = -\langle n\rangle + N\langle\beta_m\rangle - \langle\beta_m n\rangle, \tag{3}$$

where averages are taken over p_{mn}. We see that indeed the final term carries the nonlinearity. Even for the simplest coupling function, i.e. linear coupling $\beta_m = cm$, one finds a hierarchy of moment dependencies that

does not close:

$$\begin{aligned}
\partial_t\langle n\rangle &= -\gamma\langle n\rangle + \gamma cN\langle m\rangle - \gamma c\langle mn\rangle, &(4)\\
\partial_t\langle mn\rangle &= \alpha M\langle n\rangle - (\alpha+\gamma+1)\langle mn\rangle + \gamma cN\langle m^2\rangle - \gamma c\langle m^2 n\rangle, &(5)\\
\partial_t\langle m^2 n\rangle &= \ldots &(6)
\end{aligned}$$

That is, the dynamics of $\langle n\rangle$ depend on $\langle mn\rangle$, whose dynamics depend on $\langle m^2 n\rangle$, and so on.

The fact that the moments cannot be computed—indeed, not even the mean output $\langle n\rangle$—makes it particularly important to actually solve the CME in order to learn about the statistical properties of this system.

2.2 The spectrum of the switch operator

The CME is a linear equation. Even when the rates are nonlinear functions of the molecule numbers, the CME is still linear in its degree of freedom, the joint probability. The most straightforward way to solve a linear equation is to write its solution as an expansion in the eigenfunctions of the linear operator. Although it is difficult to derive the eigenfunctions of the coupled operator $\mathcal{L}_m(\alpha, M)+\gamma\mathcal{L}_n(\beta_m, N)$, it is straightforward to derive the eigenfunctions of the uncoupled operator $\mathcal{L}_m(\alpha, M)$, which we call the switch operator. Indeed, we will see that expanding the joint probability in eigenfunctions of the uncoupled operator greatly simplifies the form of the CME, yielding an exact solution in terms of matrix algebra.

The switch operator governs the CME for the first species $\mathcal{X}$; explicitly,

$$\dot{p}_m = -\mathcal{L}p_m = \alpha[M-(m-1)]p_{m-1} + (m+1)p_{m+1} - [\alpha(M-m)+m]p_m, \tag{7}$$

where for notational simplicity we have taken $\mathcal{L}_m(\alpha, M) \to \mathcal{L}$. Its eigenvalue relation is written

$$\mathcal{L}\phi_m^j = \lambda_j \phi_m^j, \tag{8}$$

for eigenvalues λ_j and eigenvectors ϕ_m^j.

2.2.1 Eigenvalues

The matrix form of the operator $\mathcal{L}$ can be read directly from Eqn. 7:

$$\mathbf{L} = \begin{pmatrix}
M\alpha & -1 & & & & & & \\
-M\alpha & (M-1)\alpha+1 & -2 & & & & & \\
 & -(M-1)\alpha & (M-2)\alpha+2 & -3 & & & & \\
 & & \ddots & \ddots & \ddots & & & \\
 & & & -3\alpha & 2\alpha+(M-2) & -(M-1) & & \\
 & & & & -2\alpha & \alpha+(M-1) & -M & \\
 & & & & & -\alpha & M &
\end{pmatrix}. \tag{9}$$

The tridiagonal structure follows from the fact that molecule numbers only increase or decrease by one at a time. Practically speaking, the eigenvalues can be obtained using the fact that the determinant of a tridiagonal matrix can be computed recursively. Performing the computation for $M = 0, 1, 2, \ldots$ reveals the pattern

$$\lambda_j = (\alpha+1)j, \qquad j \in \{0, 1, 2, \ldots, M\}. \tag{10}$$

However, Eqn. 10 can be derived more rigorously by making use of a generating function. We present this derivation next, since the generating function formalism will also prove quite useful in deriving the eigenvectors and solving the CME.

The generating function is an expansion in any complete basis for which the probability distribution provides the expansion coefficients [6]. Choosing as our basis the set of polynomials in some continuous variable x, the generating function is defined

$$G(x) = \sum_{m=0}^{M} p_m x^m. \tag{11}$$

The probability distribution is recovered via the inverse transform

$$p_m = \frac{1}{m!}\partial_x^m [G(x)]_{x=0}. \tag{12}$$

A key utility of the generating function is turning the CME, which is a set of ordinary differential equations (ODEs), into a single partial differential equation. Indeed, summing Eqn. 7 against x^m yields

$$\dot{G} = -(x-1)[(\alpha x+1)\partial_x - \alpha M]G, \tag{13}$$

where the appearances of x and ∂_x arise from the shifts $m-1$ and $m+1$, respectively. Eqn. 13 directly gives the form of the operator in x space: $\mathcal{L} = (x-1)[(\alpha x+1)\partial_x - \alpha M]$. The eigenfunctions are then obtained from the relation $\mathcal{L}\phi^j(x) = \lambda_j \phi^j(x)$ by separating variables and integrating:

$$\phi^j(x) = (\alpha+1)^{-M}(x-1)^{\lambda_j/(\alpha+1)}(\alpha x+1)^{M-\lambda_j/(\alpha+1)}. \tag{14}$$

Here, the constant factor $(\alpha+1)^{-M}$ is determined by application of the normalization condition $G(1) = 1$ to the steady state solution, which is obtained by setting $\lambda_j = 0$:

$$G(x) = \left(\frac{\alpha x+1}{\alpha+1}\right)^M. \tag{15}$$

We will solve Eqn. 13 in two ways: by the method of characteristics and by expansion in the eigenfunctions; together these solutions will reveal the eigenvalues.

First, the method of characteristics [7] posits that the dependence of G on x and t occurs via some parametric variable s, i.e. $G(x,t) = G[x(s),t(s)]$. The chain rule then gives $dG/ds = (\partial G/\partial x)(dx/ds) + (\partial G/\partial t)(dt/ds)$, which when compared term by term with Eqn. 13 yields three ordinary differential equations:

$$\frac{dt}{ds} = 1, \qquad \frac{dx}{ds} = (x-1)(\alpha x+1), \qquad \frac{dG}{ds} = \alpha M(x-1). \tag{16}$$

The first identifies $s = t$, with which the second is solved by

$$z = \frac{x-1}{\alpha x+1}e^{-(\alpha+1)t}, \tag{17}$$

where z is a constant of integration. The crux of the method is that Eqn. 17 defines a characteristic curve on which G must depend, i.e. $G(x,t) = f[z(x,t)]g(x,t)$, where f and g are unknown functions, and z has been promoted to a characteristic function of x and t. The function g is identified by realizing that steady state is reached as $t \to \infty$, for which $f(z) \to f(0)$ no longer depends on x or t. Therefore, g must be the steady state function given in Eqn. 15:

$$G(x,t) = \left(\frac{\alpha x+1}{\alpha+1}\right)^M f(z). \tag{18}$$

Although we still do not know f, we may Taylor expand it around the point $z = 0$, yielding

$$G(x,t) = \left(\frac{\alpha x+1}{\alpha+1}\right)^M \sum_{j=0}^{\infty} c_j z^j = \left(\frac{\alpha x+1}{\alpha+1}\right)^M \sum_{j=0}^{\infty} c_j \left(\frac{x-1}{\alpha x+1}\right)^j e^{-(\alpha+1)jt}, \tag{19}$$

where $c_j \equiv \partial_z^j [f(z)]_{z=0}/j!$.

Second, because Eqn. 13 is linear, we may also write down its solution as an expansion in the eigenfunctions of its linear operator:

$$G(x,t) = \sum_j C_j(t)\phi^j(x). \tag{20}$$

Under the assumption that the eigenfunctions are orthogonal (which will be shown in the next section), inserting Eqn. 20 into Eqn. 13 yields an independent ODE for each C_j, $\dot{C}_j = -\lambda_j C_j$, which is solved by $C_j(t) = c_j e^{-\lambda_j t}$ for initial conditions c_j. Inserting this functional form and that for $\phi_j(t)$ (Eqn. 14) into Eqn. 20 yields

$$G(x,t) = \left(\frac{\alpha x+1}{\alpha+1}\right)^M \sum_j c_j \left(\frac{x-1}{\alpha x+1}\right)^{\lambda_j/(\alpha+1)} e^{-\lambda_j t}. \tag{21}$$

Comparison of Eqns. 19 and 21 reveals both the expression for the eigenvalues, $\lambda_j = (\alpha+1)j$, and a limit on their domain, the nonnegative integers $j \in \{0,1,2,\ldots,\infty\}$. Of course, the domain can be a subset of the nonnegative integers; then some c_j in Eqn. 21 would be zero. Indeed, since $\mathbf{L}$ is a finite matrix of size $M+1$ by $M+1$ (Eqn. 9), it is spanned by $M+1$ linearly independent eigenvectors, meaning we expect only $M+1$ eigenvalues. In fact, the only set of $M+1$ nonnegative integers that satisfies the requirement that the trace of $\mathbf{L}$, $\sum_{m=0}^{M}[(M-m)\alpha+m] = (\alpha+1)M(M+1)/2$, equals the sum of the eigenvalues, $\sum_j(\alpha+1)j$, is $j \in \{0,1,2,\ldots,M\}$. Thus, we arrive at the result

$$\lambda_j = (\alpha+1)j, \qquad j \in \{0,1,2,\ldots,M\}, \tag{22}$$

as proposed by inspection in Eqn. 10.

2.2.2 State space notation

The linear algebraic manipulations we have done thus far can be cast in the more abstract notation of state spaces, commonly used in quantum mechanics [8]. We will find this notation useful in later sections, for example in transforming between the molecule number basis and the eigenbasis. Specifically, we introduce a state $|p\rangle$ that can be projected into $\langle m|$ space to give the probability distribution, or into $\langle x|$ space to give the generating function:

$$\langle m|p\rangle = p_m, \qquad \langle x|p\rangle = G(x). \tag{23}$$

In the same way, the jth eigenstate $|j\rangle$ is projected into $\langle m|$ space to give the jth eigenvector, or into $\langle x|$ space to give the jth eigenfunction:

$$\langle m|j\rangle = \phi^j_m, \qquad \langle x|j\rangle = \phi^j(x). \tag{24}$$

This notation offers new insight into our definition of the generating function. For example, Eqn. 11 can now be written

$$\langle x|p\rangle = \sum_{m=0}^{M} \langle x|m\rangle\langle m|p\rangle, \tag{25}$$

where we have recognized

$$\langle x|m\rangle = x^m \tag{26}$$

as the projection of the state $|m\rangle$ into $\langle x|$ space. Eqn. 25 has a clear interpretation: we have inserted a complete set of $|m\rangle$ states. Similarly, Eqn. 12 can now be written

$$\langle m|p\rangle = \oint đx \frac{G(x)}{x^{m+1}} = \oint đx \langle m|x\rangle\langle x|p\rangle. \tag{27}$$

In the first step, we have rewritten Eqn. 12 using Cauchy's theorem, where $đx \equiv dx/2\pi i$, and the contour surrounds the pole at $x = 0$. In the second step, we have recognized

$$\langle m|x\rangle = \frac{1}{x^{m+1}} \tag{28}$$

as the conjugate to $\langle x|m\rangle$. Eqn. 27 has the clear interpretation of inserting a complete set of $|x\rangle$ states, under an inner product defined by the complex integration. The choice of inner product and of conjugate state are made such that orthonormality is preserved, a fact which we may confirm by again employing Cauchy's theorem:

$$\langle m|m'\rangle = \oint đx \langle m|x\rangle\langle x|m'\rangle = \oint đx \frac{x^{m'}}{x^{m+1}} = \frac{1}{m!}\partial_x^m \left[x^{m'}\right]_{x=0} \theta(m>0) = \delta_{mm'}. \tag{29}$$

Finally, the dynamics in Eqn. 13 can be written in state space as

$$|\dot{p}\rangle = -\hat{\mathcal{L}}|p\rangle = -(\hat{a}^+ - 1)[(\alpha\hat{a}^+ + 1)\hat{a}^- - \alpha M]|p\rangle, \tag{30}$$

where we have defined the operators $\hat{a}^+$ and $\hat{a}^-$ whose projections in x space are $\langle x|\hat{a}^+ = x$ and $\langle x|\hat{a}^- = \partial_x$. These are analogous to the raising and lowering operators in the well known treatment of the quantum harmonic oscillator. This operator formalism for the generating function was first developed in the 1970s; for a review see [9].

2.2.3 Eigenvectors

The state space notation facilitates a derivation of the functional form of the eigenvectors:

$$\phi_m^j = \langle m|j\rangle = \oint đx \langle m|x\rangle\langle x|j\rangle = \oint đx \frac{1}{x^{m+1}} \frac{(x-1)^j(\alpha x+1)^{M-j}}{(\alpha+1)^M}. \tag{31}$$

Here we have inserted the eigenfunctions

$$\phi_j(x) = \langle x|j\rangle = \frac{(x-1)^j(\alpha x+1)^{M-j}}{(\alpha+1)^M} \tag{32}$$

from Eqn. 14, with eigenvalues given by Eqn. 22. We use Cauchy's theorem to perform the integration and recognize that derivatives of a product follow a binomial expansion:

$$\begin{aligned}
\phi_m^j &= \frac{1}{(\alpha+1)^M}\frac{1}{m!}\partial_x^m \left[(\alpha x+1)^{M-j}(x-1)^j\right]_{x=0} && (33)\\
&= \frac{1}{(\alpha+1)^M}\frac{1}{m!}\sum_{\ell=0}^{m}\binom{m}{\ell}\partial_x^\ell \left[(\alpha x+1)^{M-j}\right]_{x=0}\partial_x^{m-\ell}\left[(x-1)^j\right]_{x=0} && (34)\\
&= \frac{1}{(\alpha+1)^M}\frac{1}{m!}\sum_{\ell=0}^{m}\frac{m!}{(m-\ell)!\ell!}\left[\frac{(M-j)!\alpha^\ell}{(M-j-\ell)!}\theta(\ell\le M-j)\right]\left[\frac{j!(-1)^{j-m+\ell}}{(j-m+\ell)!}\theta(m-\ell\le j)\right] && (35)\\
&= \frac{(-1)^{j-m}}{(\alpha+1)^M}\sum_{\ell\in\Omega}\binom{M-j}{\ell}\binom{j}{m-\ell}(-\alpha)^\ell. && (36)
\end{aligned}$$

Here the domain Ω results from the derivatives and is defined by $\max(0, m-j) \le \ell \le \min(m, M-j)$. Eqn. 36 gives the expression for the eigenvectors. For $j = 0$ the expression reduces to the binomial distribution in terms of the occupancy $q = \alpha/(\alpha+1)$, as it must, since this is the steady state of the uncoupled process:

$$\phi_m^0 = \binom{M}{m}\frac{\alpha^m}{(\alpha+1)^M} = \binom{M}{m}q^m(1-q)^{M-m}. \tag{37}$$

This function has one maximum, and in general the jth eigenvector has $j+1$ extrema, making the eigenvectors qualitatively similar to Fourier modes or eigenfunctions of the quantum harmonic oscillator.

The switch operator $\hat{\mathcal{L}}$ is not Hermitian. A consequence is that its conjugate eigenvectors $\psi_m^j = \langle j|m\rangle$ (row vectors) are not complex conjugates of its eigenvectors $\phi_m^j = \langle m|j\rangle$ (column vectors). Rather, they

are distinct functions that must be constructed to obey an orthonormality relation in order to constitute a complete basis. The orthonormality relation can be used to derive their form in x space, $\psi^j(x) = \langle j|x\rangle$:

$$\delta_{jj'} = \langle j|j'\rangle = \oint đx \langle j|x\rangle\langle x|j'\rangle = \oint đx\ \psi^j(x)\frac{(x-1)^{j'}(\alpha x+1)^{M-j'}}{(\alpha+1)^M} = \oint đz_0\ z_0^{j'} f_j(z_0). \tag{38}$$

Here we have defined $z_0 \equiv (x-1)/(\alpha x+1)$ and $f_j(z_0) \equiv \psi^j(x)(\alpha x+1)^{M+2}/(\alpha+1)^{M+1}$ in order to draw an equivalence between Eqn. 38 and Eqn. 29, which then implies $f_j(z_0) = 1/z_0^{j+1} = (\alpha x+1)^{j+1}/(x-1)^{j+1}$, or

$$\psi^j(x) = \frac{(\alpha+1)^{M+1}}{(\alpha x+1)^{M-j+1}(x-1)^{j+1}}. \tag{39}$$

Eqn. 39 gives the form of the conjugate eigenfunctions in x space, which can be used to derive the expression for the conjugate eigenvectors as in Eqns. 31-36:

$$\begin{aligned}
\psi^j_m &= \langle j|m\rangle = \oint đx \langle j|x\rangle\langle x|m\rangle = \oint đx \frac{(\alpha+1)^{M+1}}{(\alpha x+1)^{M-j+1}(x-1)^{j+1}} x^m & (40)\\
&= \sum_{\ell\in\Omega} \binom{M-j+\ell}{\ell}\binom{m}{j-\ell}(-\alpha)^\ell(\alpha+1)^{j-\ell}. & (41)
\end{aligned}$$

Here Ω is defined by $\max(0, j-m) \le \ell \le j$. Eqn. 41 gives the expression for the conjugate eigenvectors. They are jth order polynomials in m.

2.3 Expanding the coupled problem in uncoupled eigenfunctions

We now solve the CME by expanding the solution in the eigenfunctions of the uncoupled operator. This procedure is most easily done in state space, in which the CME reads

$$|\dot{p}\rangle = -[\hat{\mathcal{L}}_x(\alpha) + \gamma\hat{\mathcal{L}}_{xy}]|p\rangle \tag{42}$$

where

$$\begin{aligned}
\hat{\mathcal{L}}_x(\alpha) &= (\hat{a}_x^+ - 1)[(\alpha\hat{a}_x^+ + 1)\hat{a}_x^- - \alpha M], & (43)\\
\hat{\mathcal{L}}_{xy} &= (\hat{a}_y^+ - 1)[(\hat{\beta}_x\hat{a}_y^+ + 1)\hat{a}_y^- - \hat{\beta}_x N], & (44)
\end{aligned}$$

as in Eqn. 30, and we have introduced the operator $\hat{\beta}_x$ whose action on the state $|m\rangle$ yields the coupling function, $\hat{\beta}_x|m\rangle = \beta_m|m\rangle$. The first step is to write the full operator as two uncoupled operators plus a correction term. Introducing the constant $\bar{\beta}$ to parameterize the second uncoupled operator, the CME becomes

$$|\dot{p}\rangle = -[\hat{\mathcal{L}}_x(\alpha) + \gamma\hat{\mathcal{L}}_y(\bar{\beta}) + \gamma\hat{\Gamma}_x\hat{\Delta}_y]|p\rangle \tag{45}$$

where we have explicitly denoted the fact that the correction term $\hat{\mathcal{L}}_{xy} - \hat{\mathcal{L}}_y(\bar{\beta})$ factorizes into two operators that act on each of the x and y sectors alone:

$$\begin{aligned}
\hat{\Gamma}_x &\equiv \hat{\beta}_x - \bar{\beta}, & (46)\\
\hat{\Delta}_y &\equiv (\hat{a}_y^+ - 1)(\hat{a}_y^+\hat{a}_y^- - N). & (47)
\end{aligned}$$

The second step is to expand the solution in the eigenfunctions of the two uncoupled operators. Introducing k as the mode index for the eigenstates of $\hat{\mathcal{L}}_y(\bar{\beta})$, we write

$$|p\rangle = \sum_{j=0}^{M}\sum_{k=0}^{N} G_{jk}|j,k\rangle. \tag{48}$$

Inserting this form into the CME, projecting with the conjugate state $\langle j', k'|$, and summing over j and k yields the dynamics for the expansion coefficients G_{jk}:

$$\dot{G}_{jk} = -[(\alpha+1)j + \gamma(\bar{\beta}+1)k]G_{jk} - \gamma \sum_{j'=0}^{M} \Gamma_{jj'} \sum_{k'=0}^{N} \Delta_{kk'} G_{j'k'}. \tag{49}$$

Here the first term is diagonal and reflects the actions of the uncoupled operators on their eigenstates. The second term contains the corrections $\Gamma_{jj'} = \langle j|\hat{\Gamma}_x|j'\rangle$ and $\Delta_{kk'} = \langle k|\hat{\Delta}_y|k'\rangle$. The first correction is directly evaluated by inserting a complete set of m states:

$$\Gamma_{jj'} = \sum_{m=0}^{M} \langle j|(\hat{\beta}_x - \bar{\beta})|m\rangle\langle m|j'\rangle = \sum_{m=0}^{M} \langle j|m\rangle(\beta_m - \bar{\beta})\langle m|j'\rangle \tag{50}$$

$$= \sum_{m=0}^{M} \psi_m^j(\beta_m - \bar{\beta})\phi_m^{j'}. \tag{51}$$

We see that $\Gamma_{jj'}$ is the simply the difference between the coupling function and the constant parameter, rotated into eigenspace. Notably, for linear coupling, $\Gamma_{jj'}$ is tridiagonal (see Sec. 2.5). The second correction is most easily evaluated by inserting a complete set of y states; the result, derived in Sec. 2.5, is

$$\Delta_{kk'} = k\delta_{kk'} - (N-k+1)\delta_{k-1,k'}. \tag{52}$$

We see that $\Delta_{kk'}$ is subdiagonal in k, which simplifies the dynamics of G_{jk} to

$$\dot{G}_{jk} = -\sum_{j=0}^{M} \Lambda_{jj'}^k G_{j'k} + \gamma(N-k+1)\sum_{j=0}^{M} \Gamma_{jj'} G_{j',k-1}, \tag{53}$$

where we define the matrix acting on the diagonal part as

$$\Lambda_{jj'}^k \equiv [(\alpha+1)j + \gamma(\bar{\beta}+1)k]\delta_{jj'} + \gamma k \Gamma_{jj'}. \tag{54}$$

The subdiagonality allows one to write the steady state of Eqn. 53 as an iterative scheme, by which the kth column of G_{jk} is computed from the $(k-1)$th column:

$$\vec{G}_k = \gamma(N-k+1)\mathbf{\Lambda}_k^{-1}\mathbf{\Gamma}\vec{G}_{k-1}. \tag{55}$$

The scheme is initialized with

$$\vec{G}_0 = \delta_{j0} \tag{56}$$

(see Sec. 2.5), and the joint distribution is recovered via

$$p_{mn} = \sum_{j=0}^{M}\sum_{k=0}^{N} G_{jk}\phi_m^j\phi_n^k, \tag{57}$$

which is the projection of Eqn. 48 into $\langle m, n|$ space.

Eqn. 57 constitutes an exact steady state solution to the CME, with G_{jk} computed iteratively via Eqns. 55 and 56, auxiliary matrices defined in Eqns. 51 and 54, and the eigenvectors given by Eqns. 36 and 41. Importantly, the computational complexity of the solution has been dramatically reduced: rather than solving the original CME (Eqn. 1), which requires inverting its operator of size $(M+1)(N+1)\times(M+1)(N+1)$, Eqn. 55 makes clear that it is only necessary to invert N smaller matrices of size $(M+1)\times(M+1)$, i.e. the matrices $\mathbf{\Lambda}_k$ for $k \in \{1, 2, \ldots, N\}$.

2.4 Exact expressions for moments

Now that we have an exact solution to the CME in terms of a spectral expansion, moments take an exact form in terms of the expansion coefficients. We thus circumvent the problem of moment closure, instead arriving at compact expressions that require only the inversion and multiplication of finite matrices via Eqn. 55.

Moments are most easily computed from the generating function, $G(x, y)$. For example, the νth moment of the output is

$$\langle n^\nu \rangle = \left[(y\partial_y)^\nu \, G(x=1,y) \right]_{y=1} . \tag{58}$$

In terms of the expansion, the generating function is $G(x,y) = \langle x, y|p\rangle = \sum_{j=0}^{M}\sum_{k=0}^{N} G_{jk}\langle x|j\rangle\langle y|k\rangle$, and using the fact that $\langle x=1|j\rangle = \delta_{j0}$ (Eqn. 32), we have

$$\langle n^\nu \rangle = \sum_{k=0}^{N} G_{0k} \left[(y\partial_y)^\nu \, \langle y|k\rangle \right]_{y=1} . \tag{59}$$

Inserting the expression for $\langle y|k\rangle$ (Eqn. 32) and defining $w \equiv \log y$, we obtain

$$\langle n^\nu \rangle = \sum_{k=0}^{N} G_{0k} \partial_w^\nu \left[\frac{(e^w-1)^k(\bar\beta e^w+1)^{N-k}}{(\bar\beta+1)^N} \right]_{w=0} . \tag{60}$$

At this point we recall that $\bar\beta$ is a constant we introduce to parameterize the expansion. The expression for the moments therefore cannot depend on $\bar\beta$: if we change $\bar\beta$, the expression in brackets changes, but the expansion coefficients G_{0k} also change, such that Eqn. 60 evaluates to the same $\bar\beta$-independent form. We are therefore free to set $\bar\beta$ to any value, and the choice $\bar\beta = 0$ makes the derivative easiest to evaluate. Thus we have

$$\langle n^\nu \rangle = \sum_{k=0}^{N} G_{0k} \partial_w^\nu \left[(e^w-1)^k \right]_{w=0} , \tag{61}$$

where it is now understood that G_{0k} is computed with $\bar\beta = 0$. Evaluating the derivative yields

$$\begin{aligned}
\langle n^\nu \rangle &= \sum_{k=0}^{N} G_{0k} \left[\sum_{\ell=1}^{\min(k,\nu)} \left\{ {\nu \atop \ell} \right\} \frac{k!}{(k-\ell)!} e^{\ell w} (e^w-1)^{k-\ell} \right]_{w=0} && (62)\\
&= \sum_{k=0}^{N} G_{0k} \sum_{\ell=1}^{\min(k,\nu)} \left\{ {\nu \atop \ell} \right\} \frac{k!}{(k-\ell)!} \delta_{k\ell} && (63)\\
&= \sum_{k=1}^{\min(\nu,N)} G_{0k} \left\{ {\nu \atop k} \right\} k! && (64)
\end{aligned}$$

in terms of the Stirling numbers of the second kind,

$$\left\{ {\nu \atop k} \right\} = \frac{1}{k!} \sum_{\ell=0}^{k} (-1)^{k-\ell} \binom{k}{\ell} \ell^\nu . \tag{65}$$

For example, the first moment, second moment, and variance are

$$\begin{aligned}
\langle n \rangle &= G_{01}, && (66)\\
\langle n^2 \rangle &= G_{01} + 2G_{02}, && (67)\\
\sigma_n^2 &= \langle n^2 \rangle - \langle n \rangle^2 = G_{01} + 2G_{02} - G_{01}^2 . && (68)
\end{aligned}$$

These are exact expressions for the moments in terms of the expansion coefficients G_{0k}, which are obtained by matrix inversion and multiplication via Eqn. 55, e.g. in Mathematica.

An informative special case is immediately revealed when $N = 1$, for which G_{02} does not exist, i.e. $\langle n\rangle = G_{01}$ and $\sigma_n^2 = G_{01} - G_{01}^2$, or

$$\sigma_n^2 = \langle n\rangle(1 - \langle n\rangle) \qquad (N = 1). \tag{69}$$

Here there is only one output molecule. The relationship between its mean activation and the associated noise must therefore obey the known result for a single binary switch, Eqn. 69.

2.5 Auxiliary calculations

Here we show that $\Gamma_{jj'}$ is tridiagonal for linear $\beta_m = cm$:

$$\begin{aligned}
\Gamma_{jj'} &= \langle j|\hat{\Gamma}_x|j'\rangle && (70)\\
&= \langle j|(c\hat{a}_x^+\hat{a}_x^- - \bar{\beta})|j'\rangle && (71)\\
&= -\bar{\beta}\delta_{jj'} + c\oint đx\langle j|x\rangle\langle x|\hat{a}_x^+\hat{a}_x^-|j'\rangle && (72)\\
&= -\bar{\beta}\delta_{jj'} + c\oint đx\langle j|x\rangle x\partial_x\langle x|j'\rangle && (73)\\
&= -\bar{\beta}\delta_{jj'} + c\oint đx\langle j|x\rangle x\partial_x\frac{(x-1)^{j'}(\alpha x+1)^{M-j'}}{(\alpha+1)^M} && (74)\\
&= -\bar{\beta}\delta_{jj'} + c\oint đx\langle j|x\rangle\frac{x}{(\alpha+1)^M}\Big[j'(x-1)^{j'-1}(\alpha x+1)^{M-j'} \\
&\qquad\qquad +(x-1)^{j'}(M-j')(\alpha x+1)^{M-j'-1}\alpha\Big] && (75)\\
&= -\bar{\beta}\delta_{jj'} + c\oint đx\langle j|x\rangle\frac{x(x-1)^{j'-1}(\alpha x+1)^{M-j'-1}}{(\alpha+1)^M}\left[j'(\alpha x+1)+(x-1)(M-j')\alpha\right] && (76)\\
&= -\bar{\beta}\delta_{jj'} + \frac{c}{\alpha+1}\oint đx\langle j|x\rangle\frac{x(x-1)^{j'-1}(\alpha x+1)^{M-j'-1}}{(\alpha+1)^M}\big\{j'(\alpha x+1)^2 \\
&\qquad\qquad +[\alpha(M-j')+j'](\alpha x+1)(x-1) \\
&\qquad\qquad +\alpha(M-j')(x-1)^2\big\} && (77)\\
&= -\bar{\beta}\delta_{jj'} + \frac{c}{\alpha+1}\oint đx\langle j|x\rangle\left\{\frac{(x-1)^{j'-1}(\alpha x+1)^{M-j'+1}}{(\alpha+1)^M}[j'] \right.\\
&\qquad\qquad +\frac{(x-1)^{j'}(\alpha x+1)^{M-j'}}{(\alpha+1)^M}[\alpha(M-j')+j'] \\
&\qquad\qquad \left.+\frac{(x-1)^{j'-1}(\alpha x+1)^{M-j'+1}}{(\alpha+1)^M}[\alpha(M-j')]\right\} && (78)\\
&= -\bar{\beta}\delta_{jj'} + \frac{c}{\alpha+1}\oint đx\langle j|x\rangle\left\{\langle x|j'-1\rangle j' + \langle x|j'\rangle\left[\alpha(M-j')+j'\right] + \langle x|j'+1\rangle\alpha(M-j')\right\} && (79)\\
&= -\bar{\beta}\delta_{jj'} + \frac{c}{\alpha+1}\left\{\langle j|j'-1\rangle j' + \langle j|j'\rangle\left[\alpha(M-j')+j'\right] + \langle j|j'+1\rangle\alpha(M-j')\right\} && (80)\\
&= \frac{cj'}{\alpha+1}\delta_{j,j'-1} + \left\{\frac{c[\alpha(M-j')+j']}{\alpha+1} - \bar{\beta}\right\}\delta_{jj'} + \frac{c\alpha(M-j')}{\alpha+1}\delta_{j,j'+1}. && (81)
\end{aligned}$$

Eqn. 71 recognizes that $\hat{\beta}_x = c\hat{a}_x^+\hat{a}_x^-$ is the operator representation of β_m (since $\hat{a}^+\hat{a}^-$ is the number operator, i.e. $\hat{a}_x^+\hat{a}_x^-|m\rangle = m|m\rangle$), and Eqn. 77 uses the algebraic fact that $x[j'(\alpha x+1)+(x-1)(M-j')\alpha](\alpha+1) = j'(\alpha x+1)^2 + [\alpha(M-j')+j'](\alpha x+1)(x-1) + \alpha(M-j')(x-1)^2$, which is straightforward to verify.

Here we derive Eqn. 52:

$$
\begin{aligned}
\Delta_{kk'} &= \langle k|\hat{\Delta}_y|k'\rangle && (82)\\
&= \langle k|(\hat{a}_y^+ - 1)(\hat{a}_y^+\hat{a}_y^- - N)|k'\rangle && (83)\\
&= \oint đy \langle k|y\rangle\langle y|(\hat{a}_y^+ - 1)(\hat{a}_y^+\hat{a}_y^- - N)|k'\rangle && (84)\\
&= \oint đy \langle k|y\rangle (y-1)(y\partial_y - N)\langle y|k'\rangle && (85)\\
&= \oint đy \langle k|y\rangle (y-1)(y\partial_y - N)\frac{(y-1)^{k'}(\bar{\beta}y+1)^{N-k'}}{(\bar{\beta}+1)^N} && (86)\\
&= \oint đy \langle k|y\rangle \frac{(y-1)}{(\bar{\beta}+1)^N}\Big[yk'(y-1)^{k'-1}(\bar{\beta}y+1)^{N-k'} + y(y-1)^{k'}(N-k')(\bar{\beta}y+1)^{N-k'-1}\bar{\beta} \\
&\qquad -N(y-1)^{k'}(\bar{\beta}y+1)^{N-k'}\Big] && (87)\\
&= \oint đy \langle k|y\rangle \frac{(y-1)^{k'}(\bar{\beta}y+1)^{N-k'-1}}{(\bar{\beta}+1)^N}\big[yk'(\bar{\beta}y+1) + y(y-1)(N-k')\bar{\beta} \\
&\qquad -N(y-1)(\bar{\beta}y+1)\big] && (88)\\
&= \oint đy \langle k|y\rangle \frac{(y-1)^{k'}(\bar{\beta}y+1)^{N-k'-1}}{(\bar{\beta}+1)^N}\big[k'(\bar{\beta}y+1) - (y-1)(N-k')\big] && (89)\\
&= \oint đy \langle k|y\rangle \left[k'\frac{(y-1)^{k'}(\bar{\beta}y+1)^{N-k'}}{(\bar{\beta}+1)^N} - (N-k')\frac{(y-1)^{k'+1}(\bar{\beta}y+1)^{N-(k'+1)}}{(\bar{\beta}+1)^N}\right] && (90)\\
&= \oint đy \langle k|y\rangle \left[k'\langle y|k'\rangle - (N-k')\langle y|k'+1\rangle\right] && (91)\\
&= k'\langle k|k'\rangle - (N-k')\langle k|k'+1\rangle && (92)\\
&= k'\delta_{kk'} - (N-k')\delta_{k,k'+1} && (93)\\
&= k\delta_{kk'} - (N-k+1)\delta_{k-1,k'}. && (94)
\end{aligned}
$$

Here we derive Eqn. 56:

$$
\begin{aligned}
\vec{G}_0 &= G_{j0} && (95)\\
&= \langle j, k=0|p\rangle && (96)\\
&= \sum_{m=0}^{M}\sum_{n=0}^{N}\langle j|m\rangle\langle k=0|n\rangle\langle m,n|p\rangle && (97)\\
&= \sum_{m=0}^{M}\sum_{n=0}^{N}\langle j|m\rangle p_{mn} && (98)\\
&= \sum_{m=0}^{M}\langle j|m\rangle p_m && (99)\\
&= \sum_{m=0}^{M}\langle j|m\rangle\langle m|j=0\rangle && (100)\\
&= \langle j|j=0\rangle && (101)\\
&= \delta_{j0}. && (102)
\end{aligned}
$$

Eqn. 98 uses Eqn. 41 to obtain $\langle k=0|n\rangle = 1$, and Eqn. 100 recognizes that p_m is the steady state of the uncoupled operator, $p_m = \phi_m^0 = \langle m|j=0\rangle$.

3 Supplementary figures

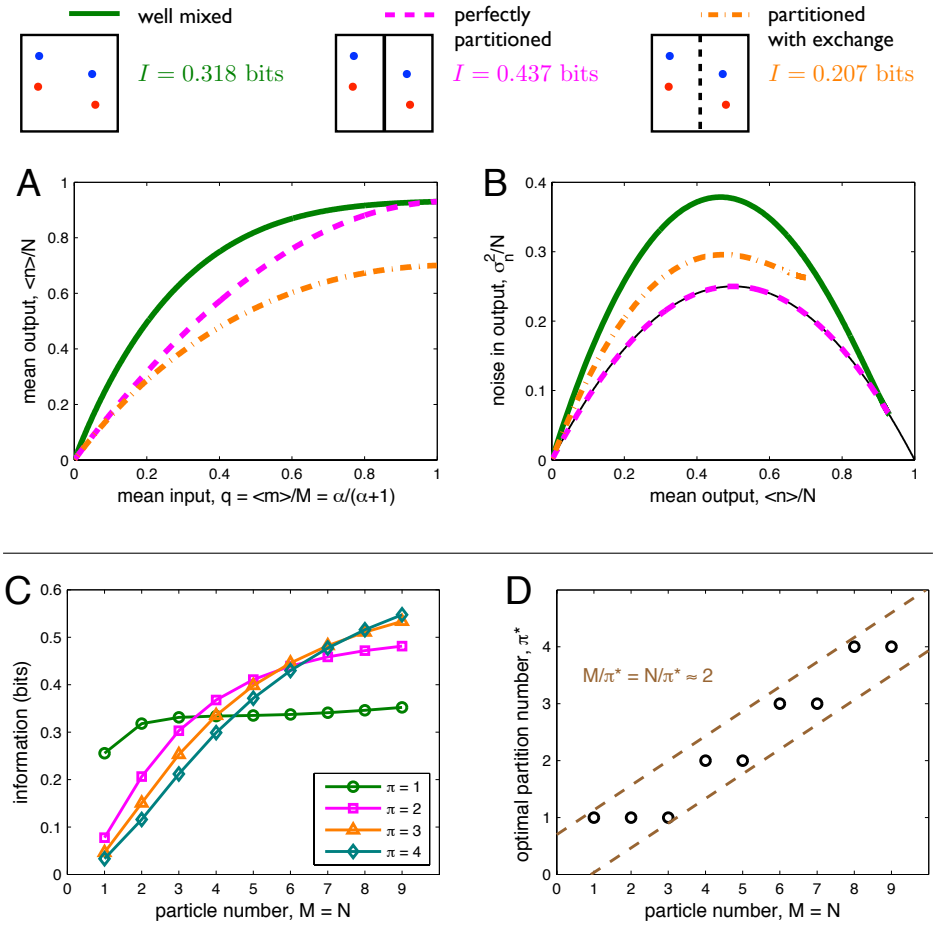

Figure S1: The effects of partitioning persist for Michaelis-Menten coupling. The coupling is described by $\beta_m^{(i)} = \beta m_i/[m_i + (V/\pi)K] = \beta m_i/(m_i + \phi M/\pi)$, where m_i is the number of X^* molecules in partition $i \in \{1, \ldots, \pi\}$, and $\phi \equiv KV/M$ is a constant. Here $\beta = 20$, $\phi = 1/2$, and $\gamma = 1$.

A, B As in Fig. 2 of the main text, with $M = N = 2$, perfect partitioning linearizes the input-output relation and reduces the noise, transmitting more information than the well-mixed system; further, allowing exchange among partitions compresses the response and increases the noise compared to the perfectly partitioned system, transmitting less information than the well-mixed system.

C, D As in Fig. 5 of the main text, an information-optimal partition size, here $M/\pi^* = N/\pi^* \approx 2$, emerges due to the trade-off between optimizing signaling reliability and avoiding unfavorable configurations.

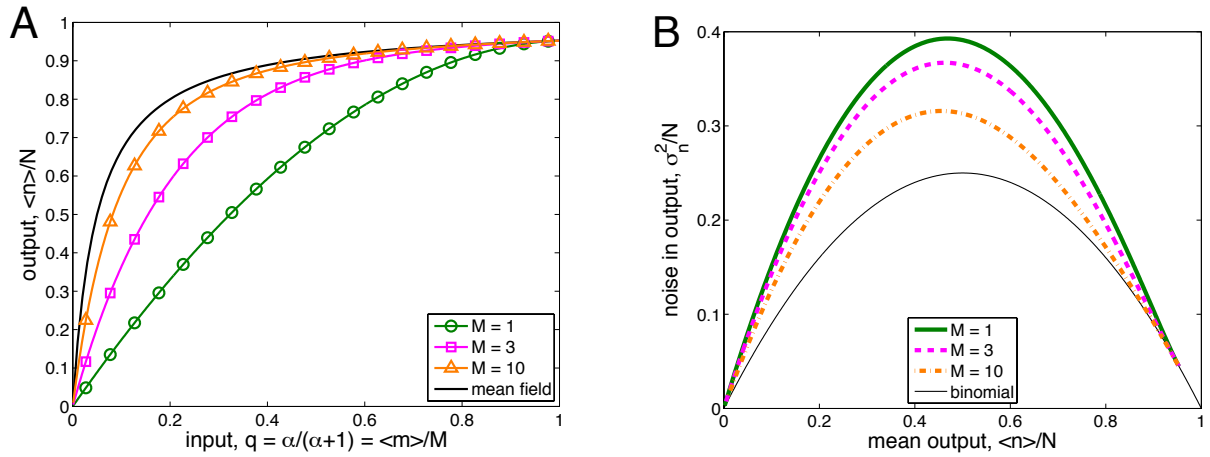

Figure S2: Reducing the number of input molecules linearizes the input-output response and increases the noise in the output. Here $\pi = 1$, $\beta = 20$, and $\gamma = 1$.

A The output (the mean activity of $N = 2$ $\mathcal{Y}$ molecules) vs. the input (the mean activity of M $\mathcal{X}$ molecules) for several values of M. As M is reduced the response becomes more linear, deviating more strongly from the mean-field response $\langle n \rangle / N = \beta q / (\beta q + 1)$. Symbols show 20 uniformly spaced values of q to highlight the effect of saturation on the state space.

B The noise vs. the mean for the output, shown for the same values of M. As M is reduced the noise increases for all values of the mean.

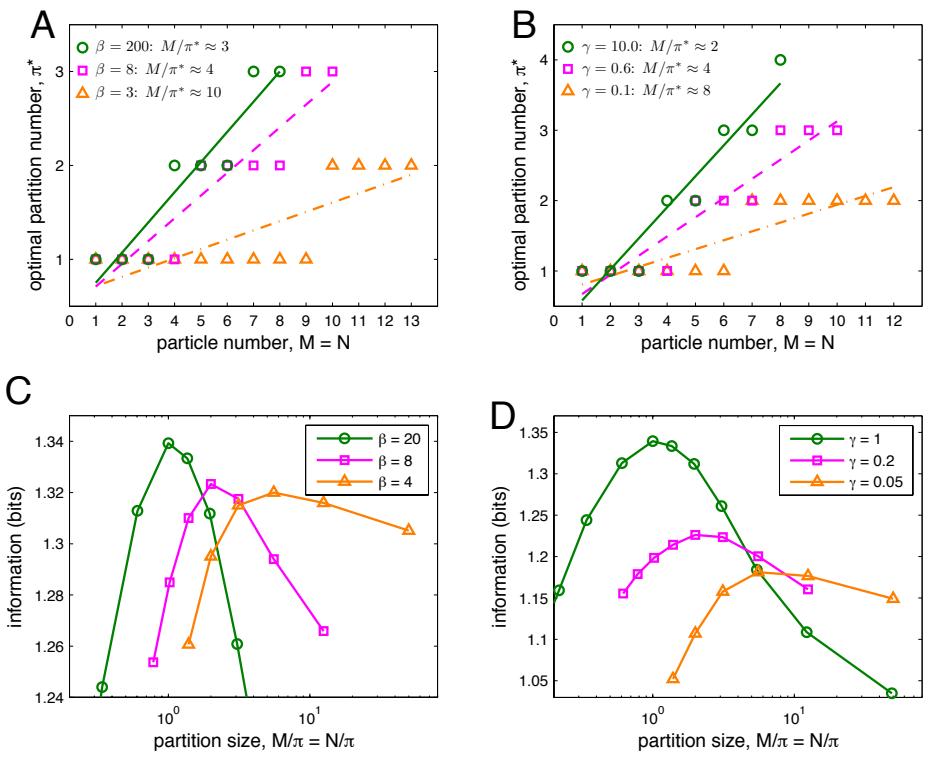

Figure S3: The emergence of an optimal partition size is robust to parameter variations.

A, B Results from the minimal system, described by the chemical master equation, as in Fig. 5B of the main text. The information-optimal partition number π^* is plotted as a function of molecule number $M = N$ for various values of β (A) and γ (B). Linear fits provide estimates of the optimal partition size M/π^*, as indicated in the legends. In A, $\gamma = 1$; in B, $\beta = 20$.

C, D Results from the lattice simulation, in which space is accounted for explicitly, as in Fig. 6C of the main text. The information is plotted as a function of the partition size, directly revealing an optimum, for various values of β (C) and γ (D). Parameters are as in Fig. 6C: $M = N = 49$, $p_{\text{hop}} = 0.001$, $\lambda = 70$, and $p_D/p_r = 1$. In C, $\gamma = 1$; in D, $\beta = 20$.

As discussed in the main text, the optimum arises due to a tradeoff between two key effects of partitioning: on the one hand, partitioning removes correlations in the states of $\mathcal{Y}$ molecules, reducing noise; on the other hand, partitioning isolates molecules, reducing the maximal response. The first effect favors few molecules per partition, while the second effect favors many molecules per partition.

As seen here in both the minimal system (A, B) and the simulated system (C, D), lowering β or γ increases the optimal number of molecules per partition. This result has an intuitive explanation in terms of the above tradeoff: lowering either β or γ slows the rate of switching from the Y to the Y^* state, with respect to the timescale of $\mathcal{X}$ switching. As a result, $\mathcal{Y}$ molecules are less sensitive to individual fluctuations in the state of $\mathcal{X}$ molecules. The states of the $\mathcal{Y}$ molecules therefore exhibit weaker correlations, which in turn weakens the benefit that partitioning imparts in terms of the removal of these correlations. The opposing effect of molecular isolation thus begins to dominate, pushing the optimum toward a larger number of molecules per partition.

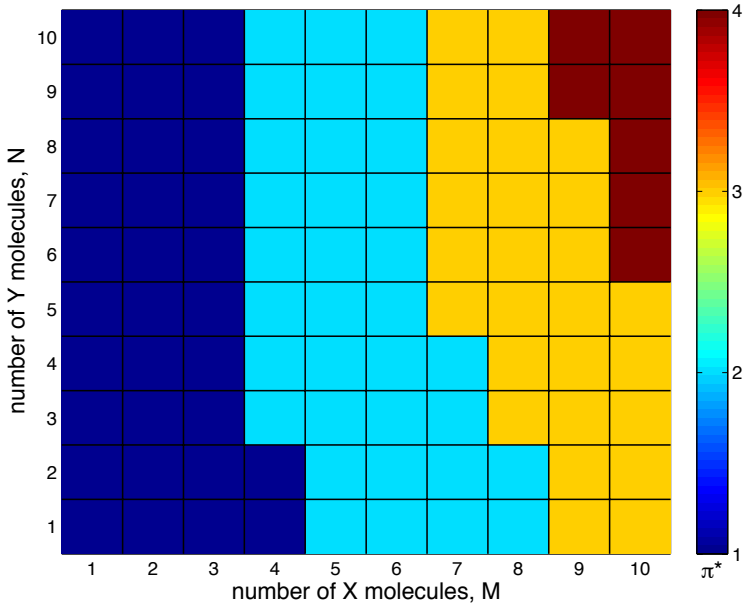

Figure S4: The optimal partition size has only a weak dependence on the number of output molecules. The information-optimal partition number π^* is plotted as a function of the number of $\mathcal{X}$ molecules M and the number of $\mathcal{Y}$ molecules N. The dependence of π^* on N is weak, such that the partition size $M/\pi^* \approx 3-4$ is roughly constant over the range of N values.

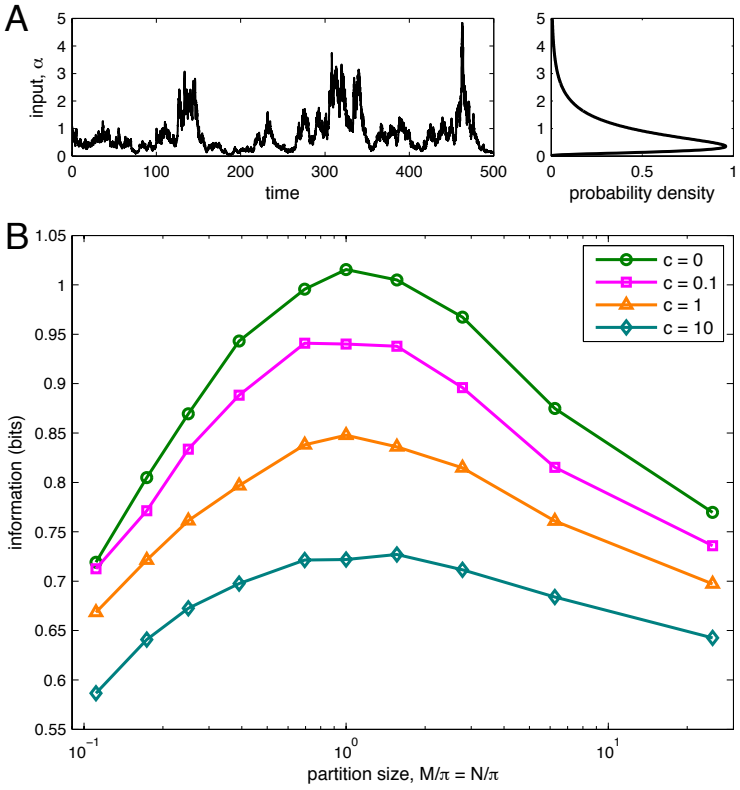

Figure S5: The effects of partitioning are robust to extrinsic noise.

A Simulations are performed with extrinsic noise introduced to the input parameter α. To keep $\alpha \geq 0$, the quantity $z \equiv \log \alpha$ is described by the simple mean-reverting Ornstein-Uhlenbeck process $dz = r(\mu - z)dt + \eta\sqrt{r dt}\xi$, where ξ is a Gaussian random variable with mean 0 and variance 1; this results in a log-normal distribution for α. The quantity $1/r$ is the autocorrelation time, and the choices $\mu = \log[\bar{\alpha}^3/(\bar{\alpha} + c)]/2$ and $\eta = \sqrt{2\log(1 + c/\bar{\alpha})}$ ensure that the mean of α is $\bar{\alpha}$ and that the variance of α scales with the mean via $\sigma_\alpha^2 = c\bar{\alpha}$.

B As the magnitude of the extrinsic noise (set by c) increases, the information $I[\bar{\alpha}, n]$ decreases for all partition sizes, while the presence of an information-optimal partition size persists.

Here $M = N = 25$, $\beta = 20$, $\gamma = 1$, $p_{\text{hop}} = 0.001$, the system is $\lambda = 50$ lattice spacings squared, the ratio of diffusion to reaction propensities is $p_D/p_r = 1$, and $r = 1$ in units of the $X^* \to X$ reaction rate (which sets the timescale of switching). In A, $\bar{\alpha} = c = 1$ and time is scaled by $1/r$. In B, when $c = 0$, the information transmission is lower than that in Fig. 6C of the main text because $M = N$ is lower.

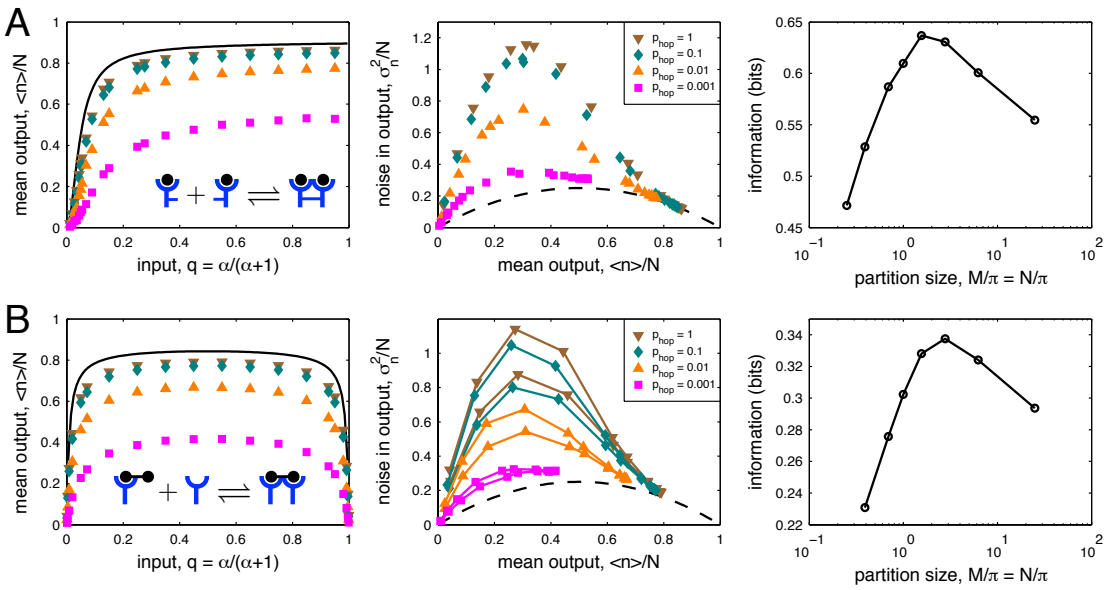

Figure S6: The effects of partitioning are robust to receptor dimerization. Two dimerization schemes are simulated, which are paradigmatic for receptor tyrosine kinases, including EGF receptor [10]: **A** Dimerization is receptor-mediated (left, inset), meaning two active receptors X^* form a complex C, or **B** dimerization is ligand-mediated (left, inset), meaning an active receptor X^* and an inactive receptor X form a complex C. The latter scheme admits a "dead-end" state at ligand saturation, when all receptors are ligand-bound and no complexes can form, leading to a non-monotonic response curve (B, left), as observed e.g. for the Ret receptor [11]. Both schemes are described by the reactions $X \underset{1}{\overset{\alpha}{\rightleftharpoons}} X^*$, $C + Y \xrightarrow{\gamma\beta} C + Y^*$, and $Y^* \xrightarrow{\gamma} Y$, with dimer formation described by $X^* + X^* \underset{\chi}{\overset{\chi\epsilon}{\rightleftharpoons}} C$ in A, or $X^* + X \underset{\chi}{\overset{\chi\epsilon}{\rightleftharpoons}} C$ in B. Here $M = N = 25$, $\beta = 20$, $\chi = \gamma = 1$, the system is $\lambda = 50$ lattice spacings squared, and the ratio of diffusion to reaction propensities is $p_D/p_r = 1$. In A, $\epsilon = 20$; in B, $\epsilon = 5$.

Left As in Fig. 6A of the main text, as the probability of crossing a diffusion barrier p_{hop} is decreased, the maximal value of the mean response decreases. In A, the response also becomes more linear, but to less of a degree than in Fig. 6A of the main text. Note that due to both finite diffusion and finite molecule number, even the unpartitioned response ($p_{\text{hop}} = 1$) deviates from the mean-field response (black solid line), which is given by $\langle n\rangle/N = \beta f/(1+\beta f)$, where f is the fraction of $\mathcal{X}$ molecules in the dimer state; in A, $f = \epsilon g^2$ with $g \equiv \langle m\rangle/M = (\sqrt{1+8\epsilon q^2} - 1)/(4\epsilon q)$, while in B, $f = \epsilon g(1-g)/(2\epsilon g + 1)$ with $g \equiv \langle m\rangle/M = [\sqrt{1+8\epsilon q(1-q)} - 1]/[4\epsilon(1-q)]$. Here $\pi = 25$. Legends in middle panels apply to left panels as well.

Middle As in Fig. 6B of the main text, as the probability of crossing a diffusion barrier p_{hop} is decreased, the output noise decreases. Black dashed line shows the binomial noise limit $\sigma_n^2/N = (\langle n\rangle/N)(1 - \langle n\rangle/N)$. In B, lines connecting data points are provided to reveal that, as there are two values of q that give the same mean $\langle n\rangle/N$ (left), the noise is higher for the smaller value of q. Here $\pi = 25$.

Right As in Fig. 6C of the main text, the tradeoff between reliable signaling (reduced noise) and efficient signaling (maintaining a high maximal response) leads to an information-optimal partition size. Here $p_{\text{hop}} = 0.001$. Here, the information transmission is lower than that in Fig. 6C of the main text because $M = N$ is lower and additionally, in B, because of the non-monotonic mean response.

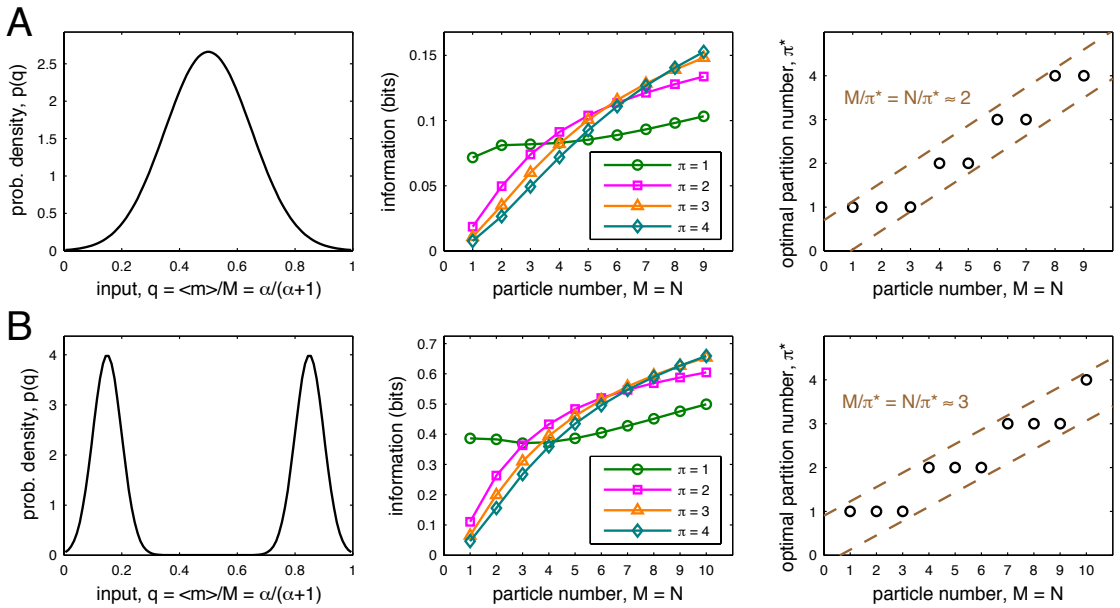

Figure S7: The effects of partitioning are robust to the shape of the input distribution. As in Fig. 5 of the main text, which takes a uniform input distribution $p(q)$, an information-optimal partition size $M/\pi^* = N/\pi^*$ persists with an input distribution that is **A** unimodal or **B** bimodal.

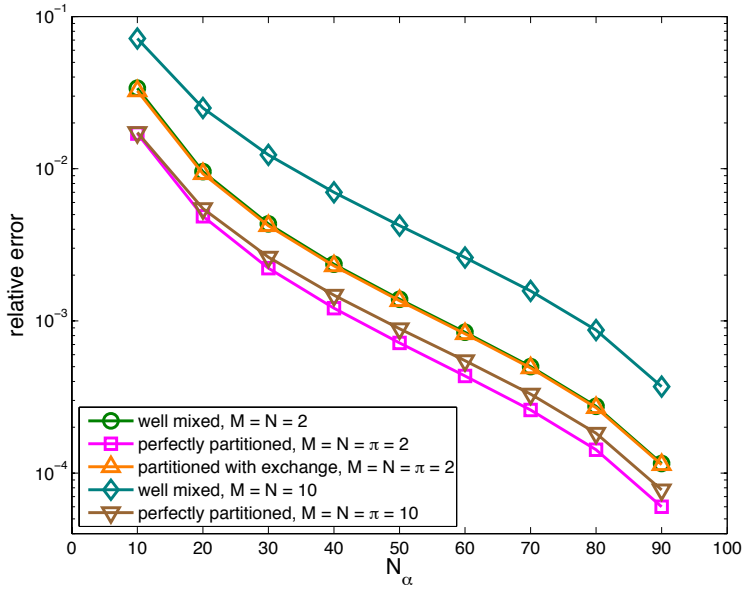

Figure S8: Computation of the mutual information converges as the input is more finely discretized. The relative error $|I - I_0|/I_0$, where I_0 is the information at $N_\alpha = 100$, is plotted against the number N_α of values of α [uniformly spaced in $q = \alpha/(\alpha + 1)$] used in the computation. Five conditions are tested, as indicated in the legend. It is seen that the relative error falls below $\sim$1% in all conditions for $N_\alpha \gtrsim 30$.

References

[1] A M Walczak, A Mugler, and C H Wiggins. A stochastic spectral analysis of transcriptional regulatory cascades. *Proc Natl Acad Sci USA*, 106:6529–6534, 2009.

[2] A Mugler, A M Walczak, and C H Wiggins. Spectral solutions to stochastic models of gene expression with bursts and regulation. *Phys Rev E*, 80:041921, 2009.

[3] C. E. Shannon. A Mathematical Theory of Communication. *Bell Syst Tech J*, 27:379–423, 623–656, 1948.

[4] A Mugler, A M Walczak, and C H Wiggins. Information-optimal transcriptional response to oscillatory driving. *Phys Rev Lett*, 105:058101, 2010.

[5] A M Walczak, A Mugler, and C H Wiggins. Analytic methods for modeling stochastic regulatory networks. In X Liu and M Betterton, editors, *Methods in Molecular Biology, Vol. 880: Computational Modeling of Signaling Networks*. Humana Press, 2012.

[6] N G van Kampen. *Stochastic processes in physics and chemistry*. Elsevier Science, 2nd edition, 1992.

[7] S J Farlow. *Partial differential equations for scientists and engineers*. Dover Publications, 1993.

[8] J S Townsend. *A modern approach to quantum mechanics*. University Science Books, 2000.

[9] D C Mattis and M L Glasser. The uses of quantum field theory in diffusion-limited reactions. *Rev Mod Phys*, 70:979–1001, 1998.

[10] M A Lemmon and J Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell*, 141:1117–34, 2010.

[11] S Schlee, P Carmillo, and A Whitty. Quantitative analysis of the activation mechanism of the multicomponent growth-factor receptor Ret. *Nat Chem Biol*, 2:636–44, 2006.